Configuration and Administration of the Unicode/Code Page Environment

Configuration and Administration of the Unicode/Code Page Environment

This chapter covers the following topics:

- ICU Library
- Profile Parameters
- Encoding Information
- Deploying Natural Objects with Encoding Information

ICU Library

Windows and UNIX Platforms

The ICU libraries are always installed with the full set of ICU conversion and collation data. The settings in the configuration file *NATCONV.INI* apply to the A format. For the U format, the corresponding checks (for example, when a character is translated to upper case) are done via the ICU library.

Mainframe Platforms

The relevant modules can be linked statically to the shared nucleus or loaded dynamically by means of the RCA technique.

Natural for Mainframes uses ICU version 3.2 which supports Unicode Version 4.0.1. Three different load modules are offered:

Load Module	Description
NATICU	Contains code page and Unicode conversion functionality as well as collation services. The first is needed for conversion from one code page to another code page or to Unicode and vice versa. The latter is used for string comparison of Unicode strings with consideration of locale ID. To keep it small, just the most popular code pages and locales are contained. The code pages are already declared in NATCONFG.
NATICUCV	Same as NATICU, but without collation services. Therefore, it is much smaller.
NATICUXL	Same as NATICU, but it contains all possible converters and locales offered by ICU 3.2. It supports about 230 different code pages (predominantly EBCDIC code pages) and 238 locales. Therefore, the module size is huge. If NATICUXL is linked to the Natural nucleus, the desired code pages have to be declared in the configuration file NATCONFG. NATICUXL supports all code pages and locale IDs which are supported by ICU 3.2 (see *http://www-950.ibm.com/software/globalization/icu/demo/converters*).

See also *NATICU for Different Purposes*.

Additional tables:

Table	Description
NATSCTU	Required scanner table for Unicode characters. It maps the properties of Unicode characters of Unicode version 4.0.1 to be used by the Natural nucleus. This table must never be changed.
NATCPTAB	Optional single-byte code page conversion accelerator tables. If the table is present, conversion from one code page to another code page will be faster since it is performed via this table rather than by calling ICU functions. For more information, see *Translation Tables* in the *Operations* documentation for Natural for Mainframes.

Profile Parameters

This section lists the profile parameters which are used in conjunction with Unicode and code page support. Not all profile parameters are available on all platforms.

- All Platforms
- Windows and UNIX Platforms
- UNIX Platforms
- Mainframe Platforms

- Natural Development Server for Mainframes

All Platforms

The following profile parameters are available on all platforms:

Parameter	Description
CP	Defines the default code page for Natural. This code page is used for the runtime and development environment if not superposed with a code page defined for a single object (for example, for a Natural source). Only platform-suitable code pages can be used. This means, for example, that no EBCDIC code page can be defined for a Windows or UNIX platform, or that no MBCS code page can be defined for a mainframe platform. On mainframes, an initialization error message occurs if a wrong code page is used. **Note:** As of Natural Version 6.2 for Windows and UNIX and Natural Version 4.2 for Mainframes, the existing CP parameter (used with Natural RPC) has been renamed to CPRPC.
CPCVERR	Specifies whether a conversion error that occurs when converting from U format to A format or from A format to U format or from between A format with different code pages results in a Natural error or not. This parameter is not regarded for the conversion of Natural sources when loading them into the source area or when cataloging them. On mainframe platforms, it is not regarded whether a Unicode field is converted into the code page before an I/O on a terminal emulation. In this case, the place holder character which is defined in NATCONFG is displayed.
CPOBJIN	Specifies the code page in which the batch input file for data is encoded. This file is defined with the Natural profile parameter CMOBJIN (Windows and UNIX) or in the dataset CMOBJIN (mainframe).
CPPRINT	Specifies the code page in which the batch output file shall be encoded. This file is defined with the Natural profile parameter CMPRINT (Windows and UNIX) or in the dataset CMPRINT (mainframe).
CPSYNIN	Specifies the code page in which the batch input file for commands is encoded. This file is defined with the Natural profile parameter CMSYNIN (Windows and UNIX) or in the dataset CMSYNIN (mainframe).
SRETAIN	Specifies that all existing sources have to be saved in their original encoding format. On Windows and UNIX platforms, the setting of the profile parameter SUTF8 has to be taken into account; see also *Customizing Your Environment*.

See also:

- *Code Pages for the Input and Output Files* in the section *Natural in Batch Mode* of the *Operations* documentation for Natural for Windows and Natural for UNIX.

- *Natural in Batch Mode* in the *Operations* documentation for Natural for Mainframes.
- For valid code pages, see *http://www.iana.org/assignments/character-sets*.

Windows and UNIX Platforms

The following profile parameter is only available on Windows and UNIX platforms:

Parameter	Description
SUTF8	Specifies the default format to be used when Natural sources are saved.

UNIX Platforms

The following profile parameter is only available on UNIX platforms:

Parameter	Description
WEBIO	Specifies whether the Web I/O Interface (which supports Unicode) or the terminal emulation window (which is not Unicode-enabled) is used for input and output.

Note:
For mainframe platforms, the NDV configuration parameter TERMINAL_EMULATION is used instead. See below.

Mainframe Platforms

The following profile parameters and/or macros are only available on mainframe platforms:

Parameter	Macro	Description
CFICU	NTCFICU	Enables Unicode support for various Unicode settings.
Not available.	NTCPAGE	In the NATCONFG module, this macro defines a code page and all related information, such as replacement characters, locale ID and collation tables.
PRINT	CP keyword subparameter of NTPRINT macro	Defines the code page for a report.
CMPO	CPAGE keyword subparameter of NTCMPO macro	Generates code page-sensitive Natural programs.
OPRB	NTOPRB	Set the ACODE and/or WCODE option to define the user encoding if the used Adabas database is enabled for UES (universal encoding support).

Natural Development Server for Mainframes

The following NDV configuration parameter is only available with the Natural Development Server for mainframe platforms:

Settings	Description
`TERMINAL_EMULATION=WEBIO`	Specifies that the Web I/O Interface (which supports Unicode) is used for input and output.

Encoding Information

The encoding of code page data can be specified on different levels:

- Level 1 - Default Code Page
- Level 2 - Code Page for a Single Object

Level 1 - Default Code Page

The default code page can be defined with the `CP` parameter. On Windows and UNIX platforms, it overwrites the system code page and is valid for all code page data.

Level 2 - Code Page for a Single Object

A code page can be defined for Natural sources, batch input (`CPOBJIN`, `CPSYNIN`) and output files (`CPPRINT`).

In addition, on Windows and UNIX platforms, a code page can be defined for work files of type ASCII, ASCII compressed, Unformatted and CSV; see *Work File Assignments* in the *Configuration Utility* documentation.

If a code page is defined at object level, this overwrites the default code page.

Note:
On Windows and UNIX platforms, it is important that the correct code page is defined for every object. For more information, see *Migrating Existing Applications*.

Deploying Natural Objects with Encoding Information

Windows and UNIX Platforms

If you want to deploy Natural objects for which encoding information has already been defined, you have to keep in mind that the encoding information is stored in the file *FILEDIR.SAG* and that it is lost if you deploy only the object file from outside of Natural.

When deploying Natural objects, you have the following possibilities for keeping the encoding information:

- You can copy the entire library. The copy of the library can then be distributed to all Windows and UNIX platforms. In this case, the original code page is kept. If a library is copied from Windows to UNIX, you have to keep in mind that it may be possible that the objects cannot be opened with a native Natural for UNIX editor because these editors can only open objects with the default code page.
- You can use the Object Handler which keeps the encoding information. In this case, the original code page is kept. If a Windows library is unloaded on UNIX, you have to keep in mind that it may be possible that the objects cannot be opened with a native Natural for UNIX editor because these editors can only open objects with the default code page.
- You can copy and paste objects with Natural Studio. In a SPoD environment, if the target environment is located on a platform different from the source environment, Natural tries to save the object with the default code page of the target environment. If this is not possible, the object is stored in UTF-8 format. For UNIX targets, this assures that the object can be opened with the native Natural for UNIX editors, if all characters of the source are available in the default code page of the UNIX server.

Mainframe Platforms

For objects on mainframe platforms, there are no special considerations for keeping the code page information of the object since it is part of the object directory.